

The separation of maximum amounts of precipitation for the Polish Atlas of Rains Intensities (PANDa)

Michał Oktawiec^{1,3,*}, *Marcin Wdowikowski*², *Bartosz Kaźmierczak*¹, *Janusz Zaleski*², and *Paweł Licznar*^{1,3}

¹Wrocław University of Science and Technology, Faculty of Environmental Engineering, Plac Grunwaldzki 13, 50-377 Wrocław, Poland

²Institute of Meteorology and Water Management – National Research Institute, Podleśna Street 61, 01-673 Warsaw, Poland

³RETENCJAPL Sp. z o.o., ul. Szymanowskiego 4, 80-280 Gdańsk, Poland

Abstract. In this paper selection of maximal precipitation amounts for Polish Atlas of Rains Intensities (PANDa) has been presented. PANDa supposes to be the source of actual and indisputable information about reliable rainfall intensities for designing of storm water drainage and retention systems in Poland. During the realization of the maximum amounts separation procedure, for chosen 100 meteorological stations with the use of peak over threshold (POT) method, a number of problems has been found, for which proceeding algorithms have been described.

1 Introduction

In Poland, the dimensioning of storm water drainage meets the basic difficulty resulting from the lack of a reliable precipitation model necessary to determine the authoritative rainfall in the national, regional or specified point range [1]. The most commonly used in engineering practice (since 1954), Błaszczyk's rainfall model underestimates the results of rainfall calculation streams by 40% [2]. This is caused by the non-stationarity (variability over the years) of rainfall – nowadays in Poland, higher sums of intense rainfalls are recorded than in the past [3–6].

Reliable precipitation models developed on the basis of current measurement data were derived only for individual towns [7]. The lack of an up-to-date atlas of rain intensities that are decisive for Poland (similar to the KOSTRA atlas in Germany) results in underestimating the required hydraulic capacity of the channels. This has its consequences when dimensioning drainage areas according to the recommendations of the European standard PN-EN 752 - directly affecting the higher incidence of urban floods in Poland [8].

The necessity of developing reliable dependencies of IDF (intensity-duration-frequency) and DDF (depth-duration-frequency) for the whole country for the purposes of designbroadly understood sanitary and water-drainage infrastructure as well as for flood

* Corresponding author: michal.oktawiec@pwr.edu.pl

protection needs has been indicated for many years by many countries. In some countries, rainfall frequency atlas exists since the 1960s. KOSTRA atlas in Germany has already a history exceeding 30 years. Therefore, Poland, using these experiences has a chance to create a tool for the 21st century [9–12].

Polish Company Retencjapl in cooperation with the Institute of Meteorology and Water Management - National Research Institute, applied to the National Center for Research and Development with a request for co-financing of the project entitled: *Development and implementation of the Polish Atlas of Rains Intensities (PANDa)* – the Smart Growth Operational Programme 2014-2020 POIR.01.01.01-00-1428/15 [13]. This article describes one of the stages of the project, which consists in separating the maximum rainfall amounts from the previously prepared (majorly digitalized from archival meteorological paper reports) rainfall time series as a partial duration series (PDS) of 16 durations from $t_1 = 5$ to $t_{16} = 4320$ min. Preparation of such an accurate numerical database of pluviographs allowed in subsequent stages to develop reliable rainfall models, which in the final stage will enable to construct a long-awaited atlas of intensive rainfall in Poland in the form of IDF and DDF dependencies. Moreover uniformly methodically developed rainfall database covering the whole country will allow to supplement the lack of knowledge about intensive rainfall in the area of interdisciplinary activities of science, climate change and civil engineering [13].

2 Materials and methods

2.1 Research area

As part of the PANDa project, a digital base of rainfall series was developed with a one-minute time step. This database, covering the years 1986–2015, consists of digitized pluviographic paper strips records (mostly rainfall data from 1986 do 2005) and registration from electronic rain gauges available from 2000 to 2015 (RG-50 SEBA working at synoptic stations and Met One Instruments 60030 - unheated and 60030H - heated, working at lower-graded IMWM-NRI stations). Rainfall data comes from a network of 100 rain gauges of the national hydrological and meteorological service, whose location is shown in Figure 1 [13].

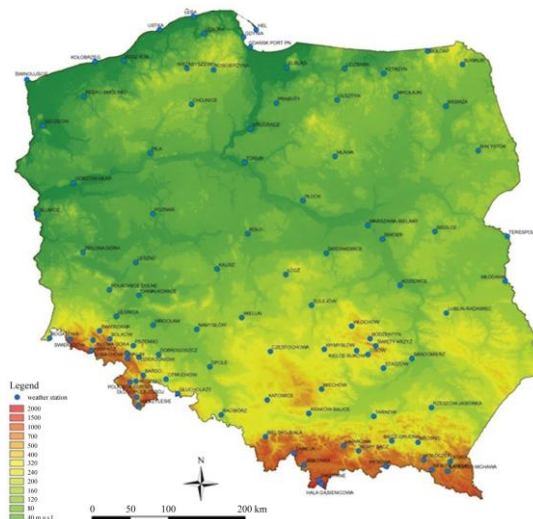


Fig. 1. Locations of IMWM-NRI rain gauges included in the digital database developed under the PANDa project (adopted from [13]).

The meteorological stations were selected for the project in such a way as to spatially cover the whole country - with the highest density in mountainous areas in the south of Poland (to better capture the orographic effect). As a result, the maximum rainfall values registered at selected 100 stations will allow to develop a rainfall atlas, which will be applied to the whole country [13].

2.2 Data selection algorithm

Input data for the calculations were one minute precipitation heights time series from 1986–2015. Primary database consisted 3000 files (30 files with a yearly series prepared as a data-value column structure, where date were stored in numerical [year-month-day hour-minute] format, for each of the 100 station) [13]. The aim of the assumed task was to isolate 30 sets of independent extreme rainfall values h_j (h_1 – the largest, ..., h_{30} – the smallest) for the 16 rainfall durations t_i : $t_1 = 5$, $t_2 = 10$, $t_3 = 15$, $t_4 = 30$, $t_5 = 45$, $t_6 = 60$, $t_7 = 90$, $t_8 = 120$, $t_9 = 180$, $t_{10} = 360$, $t_{11} = 720$, $t_{12} = 1080$, $t_{13} = 1440$, $t_{14} = 2160$, $t_{15} = 2880$ and $t_{16} = 4320$ minutes. As a result matrix of maximum rainfall partial duration series was obtained. In order to preserve the independence of maximum rainfall random variables, necessary to perform statistical analysis in the next stage of the project, measured data was isolated using complete review method. As a result 30 largest values h_j in each year (from the 30 years analyzed) were selected for each time period t_i . This approach enabled further data selection in two recommended in the literature and commonly used directions: determination of annual maximum precipitation (AMP or sometimes called annual maximum series AMS) and determination of the absolute highest precipitation values appearing above the peak-over threshold (POT).

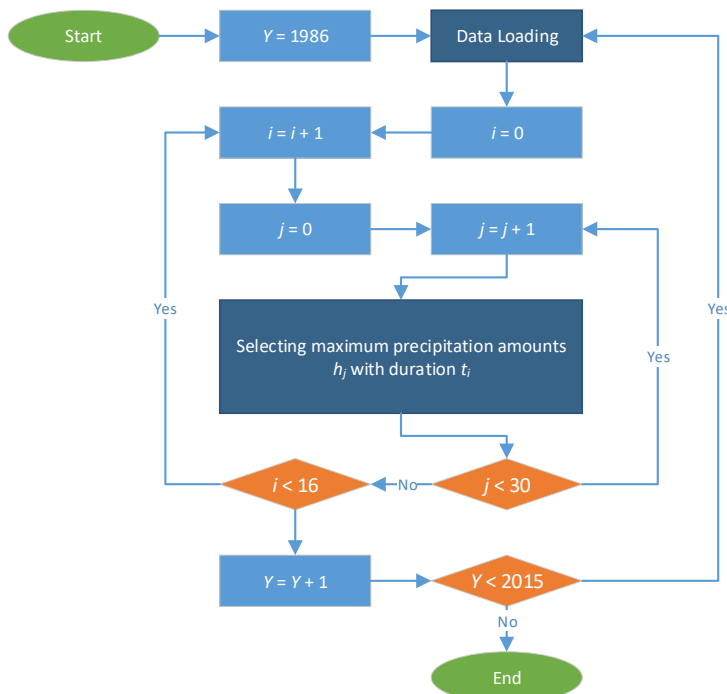


Fig. 2. Algorithm for determining the maximum rainfall partial duration series in the PANDa project.

Due to the relatively large size of the database (because of the yearly data series length, every year was in separate files), repeatability of the calculation procedure for each of the 100 stations and the obvious need to automate the calculation process, in the project self-made software based on Microsoft Excel Visual Basic for Applications (MS Excel VBA) scripts was developed. The calculation scheme was shown in the form of block diagram (Fig. 2).

The proposed algorithm (Fig. 2) for determining the maximum rainfall for a selected station in simplification presents as followed:

- step I: assigning start values ($Y = 1986, j = 1, i = 1$) and data upload to a workspace
- step II: searching for the highest independent value h_j for duration t_i then saving the result to a separate matrix and removing the workspace within the searched data. Step II internal procedure sets value $j = j + 1$, and repeats it in the loop until $j < 30$ is reached.
- step III: setting value $i = i + 1$ and return to step II. This part is repeated in the loop until $i < 16$ is reached.
- step VI: setting values $Y = Y + 1, j = 1, i = 1$ and return to step II. This part of calculations is repeated in the loop until $Y < 2015$ is reached, that is the end of the procedure at the same time.

The final results of the conducted procedure are tables containing the largest precipitation heights registered at the selected station in the years 1986–2015. For each of the 16 analyzed rainfall durations, the matrixes of the largest values (precipitation depth) in particular years as well as the matrix of absolutely the largest values in whole analyzed period (regardless to the year of occurrence) were obtained. The procedure is repeated for each of the 100 stations analyzed in the project, while the automation of the calculation process applies only to one measurement station. The following operations: summation of one-minute values in the assumed intervals of the rainfall duration, filtering and extracting 30 independent maximum values in a year, were implemented in the form of a cascade of internal procedures. However, at each stage, it was possible to run the internal procedure for one duration, e.g. only 5 or 1440 minutes.

As mentioned at the beginning of the chapter 2.2, the basic data files with one-minute rainfall amounts for each year were prepared in a time series, therefore part of the internal procedures of the program concerned calculations on time units. Due to the methodically required need to ensure the independence of individual precipitation heights with respect to each other, special attention have been paid to the mechanism identifying that kind of situation, especially for short rainfall durations, for t from 5 to 30 minutes, when during one rainfall event, e.g. 2 hours, values that fulfill the selection criteria could be from a few to a dozen. In the case of data selection using the annual maxima method, this is of no great importance, because only one value from a given year is used for further statistical analyzes, while the POT method requires caution in the data selection [14, 15].

Regarding to the independent maximum rainfall values searching procedures, it is worth to mention that a number of individual annual series records was zero, especially in years with deep meteorological drought. Such a situation did not cause methodical problems, however, it influenced the time of the automatic data search procedure. These and other problems were described in the section below.

3 Specific problems and solution proposals

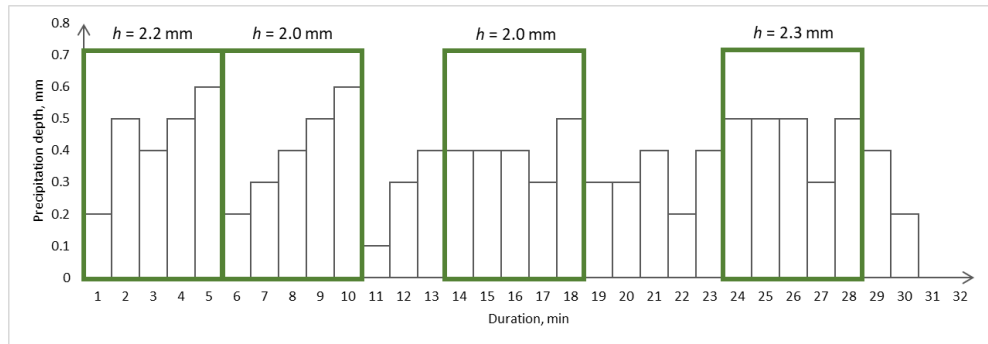
During the implementation of the separation of maximum values procedure, a few of specific problems were encountered, among which the most important were:

- What to do if at least two time-dependent precipitation depths had an equal value? Which one or maybe both choose for further analysis?

- What to do if the next largest value of rainfall was between previously selected rainfall with duration t_i , however, the length of the selected series was smaller than t_i . Should such a value be taken for further analysis?
- How to maintain the independence of rainfall data?

The first problem was exemplified in Figure 3 which presents that the largest precipitation depth $h_1 = 2.3$ mm with duration $t_1 = 5$ min occurred from 24 to 28 minutes inclusive. Another maximum height of $h_2 = 2.2$ mm, occurred from 1 to 5 minute (Fig 3 – case I), or from 2 to 6 minute (Fig. 3 – case II). Because $h_2 = 2.2$ mm in both cases was time-dependent, one of it should be chosen for further analysis. Choosing the first option, we obtained the following results: $h_1 = 2.3$ mm, $h_2 = 2.2$ mm, $h_3 = 2.0$ mm and $h_4 = 2.0$ mm, while for the second option was: $h_1 = 2.3$ mm, $h_2 = 2.2$ mm, $h_3 = 2.0$ mm and $h_4 = 1.9$ mm.

Case I



Case II

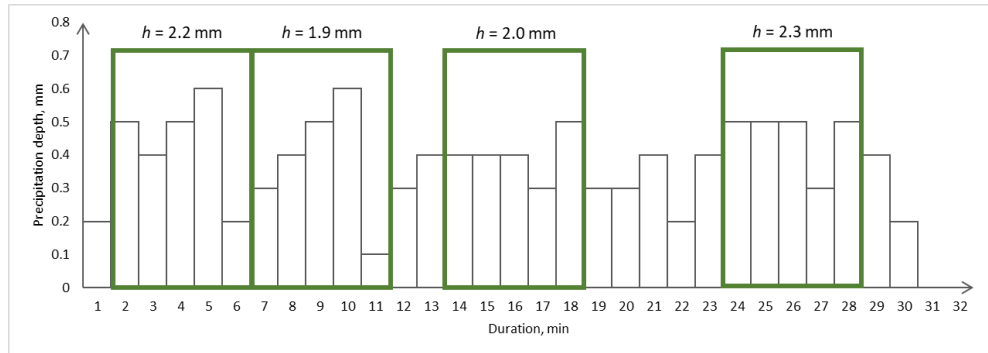


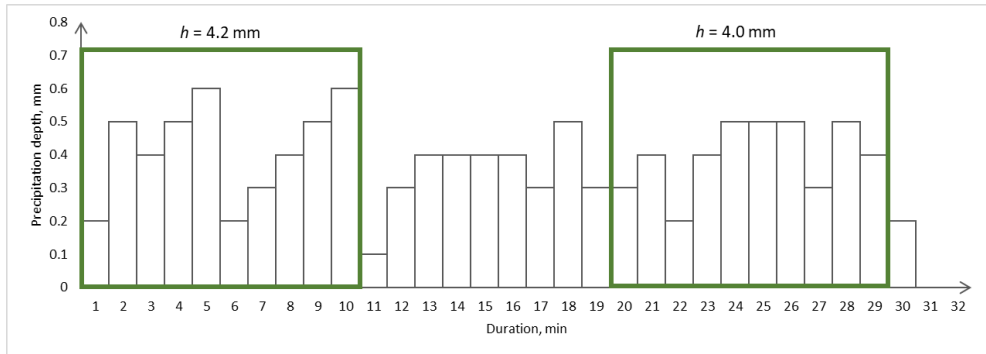
Fig. 3. Determination of maximum values with duration $t_1 = 5$ min.

Because the task is to select the maximum values, we should choose the first option. The solution to this problem required programming intervention, which consisted of:

- In the first step, the precipitation values beginning as the earliest (from time-dependent precipitation of the same height) was selected, and then the remaining precipitation sums (time-independent) were counted and ranked not descending.
- In the second step, the precipitation beginning as the latest (from time-dependent precipitation of the same height) was selected, and then the remaining precipitation sums (time-independent) were counted and ranked descending.
- In the third (final) step, both of the maximum value sequences determined in the first and second steps were compared and selected for further analysis as data string with larger values.

The second problem was exemplified in Fig. 4. The largest rainfall amounts with duration $t_2 = 10$ min were $h_1 = 4.2$ mm and $h_2 = 4.0$ mm and occurred respectively from 1 to 10 and from 20 to 29 minutes (Fig. 4 – case I). Another maximum height, amounting to $h_3 = 3.1$ mm, occurred from 11 to 19 minute – a 9-minute long string (Fig. 4 – case II).

Case I



Case II

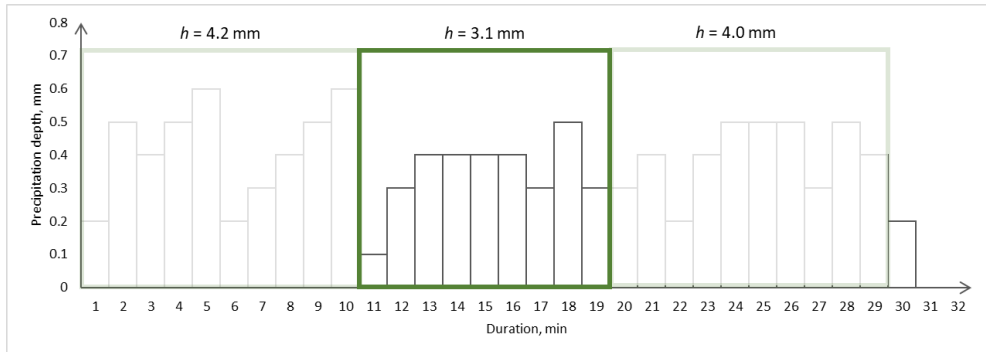


Fig. 4. Determination of precipitation maxima with duration $t_2 = 10$ min.

The second problem required making a decision whether to include rainfall in the analysis described above. After careful analysis, it was decided that a safer solution would be taken into account it in the development of rainfall formulas (rainfall height dependent on rainfall duration t and probability of exceedance p), because it was included in the sum of the 30 largest values with a given duration t .

The issue of independence of the separated interval values of precipitation in the examined cases $t = 5$ minutes (Fig. 3) and $t = 10$ minutes (Fig. 4) was also the subject of major problems. The literature does not provide clear guidelines on how to maintain the independence of rainfall maxima selecting partial duration series from 5 to 4320 minute, indicating that the safest way is to use the method of annual maxima. At the beginning, the calculation procedure assumed extracting one interval maximum value from one precipitation episode, assuming that the interval between successive precipitation is 1 day. It quickly turned out that for the rainfall durations from three hours to three days, it was not possible to complete the 30 largest values in a year that fulfilled the accepted POT criterion. Large precipitation values of short duration were also omitted, which occurred for example at the beginning and then at end of the precipitation episode. In the next approach, an analysis was performed in which the interval between the values might have no exceeded the length of the analyzed range of precipitation duration, as in Figure 4. The results were much better but for several durations still insufficient, that is why in view of the time-

consuming calculations, it was considered to accept data selection from successive intervals as a final approach. The final control of data independence was implemented during the final selection of the 30 largest values from the entire multiannual period to statistical analyzes. If in the sequence of consecutive rainfall depths the selected values would come from almost the same dates, such record should have been removed, and in this place, another separated value was passed. Despite the use of quite mild criteria for the selection of maximum rainfall, a 30-year series of measurements allowed to create such a large data set that in the end all maxima were independent in time and most often came from various genetic rainfall episodes.

4 Instead of a conclusions

Both global progressive urbanization and climate change [6] have a negative impact on the efficiency of the functioning of sewerage systems, causing more and more frequent overloads leading to local flooding or urban flooding [16, 17]. In Poland, an atlas of rains intensities (PANDa) will be created soon. It will surely become a reliable source of the knowledge used in design flood control facilities and in development of risk maps, and also for sizing stormwater drainage systems whose task is to provide the required land drainage standard [13]. The rigorous requirements for the choice of random variables, the quality of which consequently determines the quality of probabilistic models, obliges the researcher to act carefully and precisely. The aim of this paper was to present the calculation algorithms for maximum rainfall interval sums selection, which were used for the project purposes. In addition, several main technical and methodological problems were presented, which the research team encountered during the development of rainfall data, for further statistical analysis. During the preparation and development of rainfall data, it was confirmed that such analyzes should be performed using the complete review method (preferably based on one-minute data). This method is much more time-consuming but it allows to identify much larger values than in case of data processing in rigid time intervals, e.g. for $t = 5$ minutes, between 10:00 and 10:04, and also between 10:05 and 10:09 etc. In the calculation procedures, very restrictive method of selecting time-independent data was not adopted for the project's needs, whereas in the final result, no values were dependent on each other.

The key problems have been solved empirically, with expert methods, whenever the accessible Polish and foreign literature sources did not indicate transparent answers.

Acknowledgements

The project entitled: Development and implementation of the Polish Atlas of Rains Intensities (PANDa) – POIR.01.01.01-001428/15, is financed by the National Centre for Research and Development under the Smart Growth Operational Programme 2014-2020, Priority axis: Support for R&D in enterprises; Measure: R&D projects in enterprises; Sub-measure: Industrial research and development conducted by enterprises

References

1. A. Kotowski, *Podstawy bezpiecznego wymiarowania odwodnień terenów* (Seidel-Przywecki, Warsaw, 2015)
2. P. Licznar, K. Siekanowicz-Grochowina, M. Oktawiec, A. Kotowski, E. Burszta-Adamiak, *Ochrona Środowiska* **40**, No. 2, 17–22 (2018)
3. A. Dai, *WIREs Climate Chang* **2**, No. 1, 45–65 (2011)

4. A.K. Fleig, L.M. Tallaksen, P. James, H. Hisdal, K. Stahl, *Hydrology and Earth System Sciences* **19**, 3093–3107 (2015)
5. Z.W. Kundzewicz, S. Kanae, S.I. Seneviratne, J. Handmer, N. Nicholls, P. Peduzzi, R. Mechler, L.M. Bouwer, N. Arnell, K. Mach, R. Muir-Wood, R. Brakenridge, W. Kron, G. Benito, Y. Honda, K. Takahashi, B. Sherstyukov, *Hydrological Sciences Journal* **59**, No. 1 (2012)
6. Q. Schiermeier, *Nature* **470**, No. 7334, 316 (2011)
7. A. Kotowski, B. Kaźmierczak, *Journal of Hydrometeorology* **14**, 1958–1965 (2013)
8. B. Kaźmierczak, A. Kotowski, *Theoretical and Applied Climatology* **118**, No. 1, 285–296 (2014)
9. S. Perica, D. Martin, S. Pavlovic, I. Roy, M.St. Laurent, C. Trypaluk, D. Unruh, M. Yekta, G. Bonnin, *NOAA Precipitation-Frequency Atlas of the United States* (Silver Spring, Maryland, 2013)
10. W.H. Asquith, M.C. Roussel, *Atlas of Depth-Duration Frequency of Precipitation Annual Maxima for Texas* (U.S. Geological Survey, Reston, Virginia, 2004)
11. G. Malitz, H. Ertel, *KOSTRA-DWD-2010, Starkniederschlagshöhen für Deutschland (Bezugszeitraum 1951 bis 2010)*, (Deutscher Wetterdienst, Offenbach am Main, 2015)
12. A.V. Dyrørdal, *Estimation of extreme precipitation in Norway and a summary of the state-of-the-art*, (Norwegian Meteorological Institute, Oslo, 2012)
13. E. Burszta-Adamiak, P. Licznar, J. Zaleski, *Meteorology, Hydrology and Water Management-Research and Operational Applications* **7** (1), 3–13 (2019)
14. S.G. Coles, *An Introduction to Statistical Modeling of Extreme values* (Springer Series in Statistics, London, 2001)
15. D. Koutsoyiannis, *Hydrological Sciences*, **49** (4), 575–590 (2004)
16. N. Bezak, M. Šraj, M. Mikoš, *Journal of Hydrology* **541**, 272–284 (2016)
17. M.A.M. Saboia, F.A. Souza Filho, L.M. Araujo Junior, C.S. Silveira, *Brazilian Journal of Water Resources* **22**, No. 21, 1–15 (2017)